Using Haptics Interaction in Bioinformatics Application

Arthurine Breckenridge
arbreck@sandia.gov

Ben Hamlet
bhamlet@nmt.edu

Sandia National Laboratories, New Mexico is continuing to research 3D user interfaces specifically with the addition of haptics, the sense of touch. Our project is researching computational and collaboration tools independent of content area. However, applying the work to specific problems is very beneficial guide to our development. The application specific content focus of this paper is a bioinformatics application.

Bioinformatics is the combination of computer science and biology (the youngest of the natural sciences). Sequence-structure-function prediction refers to the idea that, given a molecule's sequence identity, we would like to predict its three-dimensional structure and, from that structure, infer its molecular function. Researchers have uncovered reasonable evidence indicating that a protein's structure approximately determines its molecular functions, such as catalysis, DNA binding, and cell component binding.

Many results in experimental biology first appear in image form – a photo of an organism, cells, gels, or micro-array scans. As the quantity of these results accelerates, automatic extraction features and meaning from experimental images becomes critical. At the other end of the data pipeline, naïve 2D or 3D visualizations alone are inadequate for exploring bioinformatics data. Biologists need a visual environment that facilitates exploring high-dimensional data dependent on many parameters. Therefore, visualization of bioinformatics is a vast field. Our specific focus is highlighted in special interest box: intron/exon splicing.

Basic Simulation Elements

The two basic elements of a computer haptics simulation are graphics and force that actually work independently of each other. This divides the simulation into two distinct parts, each of which present unique problems that can be addressed in independent and different ways. Although a number of factors go into making a successful simulation, the success of ones application will ultimately rely on good refresh rates (i.e., +30 Hz for graphics and 1000 Hz for force feedback). Both graphics and haptics interaction can apply to the application content and the graphic/haptic user interface. The application will use e-Touch ™ originally developed within Sandia National Laboratories, New Mexico and now licensed by Novint Technologies, Inc.

Graphic/Haptic User Interface

Force feedback has been implemented into the e-Touch ™ user interface. The basic elements are documented on the www.novint.com website. Features such as craft navigation and menus with touch sensations are standard. In addition, this application has added 3D box selection of elements and magnetic pointer to start location. The start location pointer can be used as intelligent alignment tool.

Graphics Generation

A number of approaches are available for rendering molecular graphics. One is a simple shape like line, sphere, etc. Although simple to code, as the number of primitives increase, the graphics refresh rate decreases substantially. Quick fixes are to sort the data by atom type to avoid graphics transition states like color and material properties. Keep in mind that data pre-processing is highly desirable for all haptics simulations due to the relentless refresh rates that such an application demands. Another approach is to calculate which surfaces are hidden. Although removing hidden surfaces is substantially

faster than drawing them, when dealing with a large number of atoms, it can still exceed the haptics servo loop requirements. Of course, the classic solution is to draw different levels of details based on screen position, user selection via graphical user interface, and spatial decomposition

Another graphical method used that seems promising is to fool the eye that it is seeing a 3D object since the sense of touch sends another signal that the object is actually has

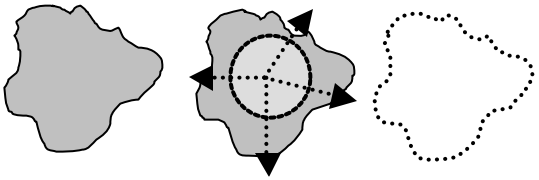

Figure 1. Circular Vector Field

shape (i.e., round atom). In this method, one determines the shape and contour of the protein's surface within a given level of detail and creates a graphics triangular surface mesh. For example, knowing the center location of each atom, the furthest sphere intercept along these vectors can be calculated (Figure 1).

Each point of intersection will then be used as a vertex in creating a surface mesh of triangles. The surface mesh can be rendered on a PC cluster since it requires an entire program to simply plot it. Once the coordinates of the surface points have been calculated, there is still a substantial amount of compute cycles needed to create a texture map for the protein and define the shadowing of the polygons. If it seems complex, this graphics solution is actually beneficial to maintain the needed haptics refresh rates and scale linearly.

Force Calculations

Computer haptics using the PHANToM is basically surface forces that act like springs, where the force exerted is equal to a constant times the depth of penetration, F= kx. Because force magnitude is dependent on the depth of penetration, when one touches two spheres at the same time but does not penetrate their surfaces an equal distance, one surface will exert a greater force than the other (Figure 2).

All of this takes place at 1000 Hz, which causes a high frequency oscillation of the PHANToM that can be felt or even heard in many circumstances. Force buzzing is simply where the PHANToM vibrates because of an inconsistent or unstable stream of forces is being fed to it. Object springiness can cause this because when two identical spheres push on the cursor, and the first sphere pushes harder than the second, the cursor will penetrate the second sphere more deeply, inducing a strong force in that sphere and so on. There are a number of solutions to this problem, many of which are not simple and may have undesirable side effects. The easiest way to eliminate buzzing, at least to the extend of audible recognition, is to vary the spring constant of the surface forces proportionally to the number of objects being touched. For example, if one begins to touch one atom, you will feel the maximum force. Then if one moves the cursor to touch more atoms, the force will decrease (.05 to .1) and will increase as you touch less. If one's range of forces and the speed at which one ramps them is reasonable, this method can alleviate solid surface buzzing.

Another solution to force buzzing produced by overlapping objects is to not have overlapping objects. Much in the same way that a graphics mesh of a protein molecule can be generated, a force shell for a protein can be created. By pre-processing the normals, one no longer has

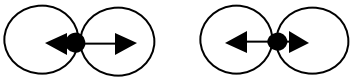

Figure 2. Force Buzzing

to worry about inter-object forces. A force shell will have the affect of making your protein completely impenetrable and may greatly reduce the value of one's simulation. Whereas the graphics shell still seemed 3D due to the force representation, this feature is lost if both graphics and force shells are used.

Special Interest: Intron/Exon splicing

The creation of life begins with a single cell that divides into two. And, why does this process sometimes malfunction, leading to defects, cancers, and other diseases? And, could this process sometimes function as planned and could potentially be used by terrorists. At the brink of the twenty-first century, there are 24 complete "draft" genomes available in public databases including that of the human genome. Though impressive information, the real challenge is transforming the torrent of raw data into biological knowledge. Thus, bioinformatics, the combination of biology and computer science is introduced. When asked what is the Holy Grail of bioinformatics, most researchers would answer the ultimate goal of a genome project is to determine the function and biological role for all of the genes of interest ideally in silico.

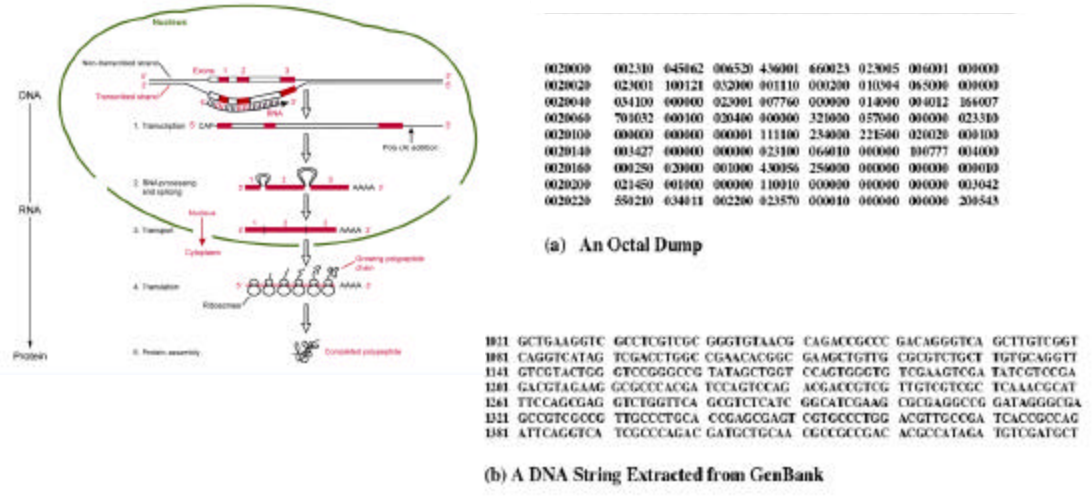

Figure A. Central Dogma as seen by a Biologist on the Right and a Computer Scientist on the Left

Though considerable progress has been made in understanding flow of information known as the "central dogma, Biology is the youngest of the natural sciences. Figure A.a) shows an octal dump of an assembly language code, which once upon a time ubiquitous in debugging computer programs, which nowadays become a rarity. Figure A.b) show a portion of the GenBank database presenting parts of DNA containing the letters A,C,G,T suggesting a quaternary kind of dump. Hopefully, this important but initial analysis will also become a rarity, as we understand more structural dynamics.

Because of the interdependent flow of information represented by the central dogma, one can begin discussion of the molecular expression of genetics of gene expression at any of its three informational levels: DNA, RNA, or protein. In 1953, James Watson and Francis Crick deduced the secondary structure of DNA. This was one of the most important biological advances, since it led to an understanding of the relationship of the DNA structure to its function, particularly to the way it was replicated. Scientists now are beginning to understand the process of copying DNA into RNA called transcription. The perceived role of RNA has changed from a passive messenger of information and scaffold for proteins to a central and active role in the functioning of the cell. To understand how a specific RNA molecule operates, its functional structure need to be better understood. It may be deduced from (costly and difficult to obtain) X-ray diffraction or NMR data only for short RNAs. In most cases, however, only the single RNA sequence (the primary structure) without further information regarding its functional form is available.

The secondary structure differs from the Watson-Crick DNA secondary structure in that it is generally single-stranded. When left in its environment, this molecule will fold itself into its secondary structure by creating base-pairs (an A with a U, a C with a G, or even a U with a G). Scientists are now deducing the secondary structure of RNA by using difference equations and dynamic programming. In 1977, it was determined that genes of higher organisms did not follow the simplest "central dogma" model. Instead, few genes exist as continuous coding sequences. Rather, one or more non-coding regions interrupt the vast majority of genes. These intervening sequences, called introns, are initially transcribed into pre-mRNA in the nucleus but are not present in the mature mRNA in the cytoplasm. Introns alternate with coding sequences, or exons, that ultimately encodes the amino acid sequence of the proteins. The portions corresponding to introns are removed, and the segments corresponding to exons are spliced together. The production of mRNA has to occur in the correct amount, in the correct place, and at the correct time during development or during the cell cycle. Each of the steps in this complex pathway is prone to error, and mutations that interfere with the individual steps have been implicated in a number of inherited genetic disorders.

Considering human genes contain ten times more intronic than exonic sequence, this weakness in the understanding of higher eukaryotes genetics is becoming increasingly apparent. Recent bioinformatic studies suggest that at least one third of human genes are alternatively spliced. Intron's evolutionary functionality is in much debate. However, there is agreement that the splicing sequences are neither strong nor unique enough signals since the splice sequence can be found in other parts of the mRNA. In the context of computer prediction of exon boundaries based only on primary sequences, this makes the task quite difficult. Therefore, what factors in the nucleus may also facilitate identification of sites? Shown in Figure A, this proposal is interested in process for detecting nuclear introns. Soon after discovery of splicing it became evident that this process also occurred in lower eukaryotes like yeast. Because the chemical mechanism of splicing and many factors involved in the splicing process are conserved from yeast to man, the budding yeast, Saccharomyces cerevisiae, has become an important model systems for the analysis of splicing. A comprehensive study of splicing in vivo of yeast (Spingola, 1999) concluded that results show that correct prediction of introns remains a significant barrier to understanding the structure, function and coding capacity of eukaryotic genomes, even in a supposedly simple system like yeast. As complete eukaryotic genome sequences become available, better methods for predicting RNA splicing signals in raw sequence will be necessary in order to discover genes and predict their expression.

Why study secondary structure? One of the major problems facing computational biology is the fact that sequence information is available in far greater quantities than information about the three-dimensional structure. While the prediction of 3D RNA structures is unfeasible at present, the prediction of secondary structures is in principle tractable. In RNA the secondary structure elements are significantly more stable and form faster than the tertiary interactions. Thus, a separation of an RNA folding model into secondary (properly nested base pairs), and tertiary (non-planar nucleotide contacts) seems feasible. Determining the secondary structure of an RNA molecule is widely seen as a first step towards understanding its biological function.

Spingola, M, Grate, L., Haussler, D., and Ares, M., "Genome-wide bioinformatic and molecular analysis of introns in Saccharomyces cerevisiae", *RNA*, 5:221-234, 1999.